DOCUMENT RESUME

ED 408 338 CS 213 831

AUTHOR Slavin, Robert E.; And Others
TITLE Effects of Full and Partial Implementations of Success for All on Student Reading Achievement in English and Spanish.
INSTITUTION Johns Hopkins Univ., Baltimore, MD. Center for Social Organization of Schools.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.; Carnegie Corp. of New York, N.Y.
PUB DATE Mar 97
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
CONTRACT R2-17-R4001S
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Early Intervention; Grade 1; *High Risk Students; Language of Instruction; Primary Education; Program Effectiveness; Program Implementation; *Reading Achievement; Reading Improvement; Reading Research; Spanish; *Tutoring
IDENTIFIERS Houston Independent School District TX; *Success for All Program

ABSTRACT

A quasi-experimental study compared achievement effects of various configurations of the Success for All (SFA), schoolwide reform model for elementary schools serving many at-risk students, from full implementation (certified tutoring, full-time school-level facilitator, and the family support team) to implementations lacking some elements felt to be essential. Full implementation of both the English and Spanish configurations of SFA in the Houston Independent School District (Texas) was achieved during the second year of the program, 1995-96. Two student cohorts participated in the English program--335 students who began first grade in 1995 and 682 students who began first grade in 1996. Subjects were in 44 SFA schools and 18 comparison schools. The final sample of subjects in the Spanish program consisted of 270 first-grade students in 21 SFA and 10 comparison schools. The Language Assessment Scales and a battery of 4 reading posttests were administered. Results indicated that: (1) programs with part-time facilitators, certified tutoring, and a family support team produced positive effects in moderately impoverished schools, but not in extremely impoverished schools; (2) schools that did not make the basic organizational changes required to provide family support and certified tutoring performed worse than comparison schools; (3) a full-time facilitator was essential in extremely high-poverty schools serving predominantly African American student populations; and (4) schools that provided certified tutoring achieved better results in the English language program. Findings suggest that schools should undertake implementation of the whole program, and that comprehensive changes in teachers' daily practices and in school organization are needed to bring about major changes. Contains 16 references, 6 tables and 3 figures of data.) (RS)

ED 408 558

Effects of Full and Partial Implementations of Success for All on Student Reading Achievement in English and Spanish

John Nunnery
Robert E. Slavin
Nancy A. Madden
Johns Hopkins University

Steven Ross
Lana Smith
University of Memphis

Phyllis Hunter
Joseph Stubbs
Houston Independent School District

PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

J. Nunnery

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Paper presented at the annual meeting of the American Educational Research Association, Chicago, March, 1997.

This research was supported by grants from the Office of Educational Research and Improvement, U.S. Department of Education (No. R-117-D40005) and the Carnegie Corporation of New York. However, any opinions expressed do not necessarily represent the positions or policies of our funders.

CS012831

BEST COPY AVAILABLE

Success for All (Slavin, Madden, Dolan, & Wasik, 1996) is a schoolwide reform model primarily for elementary schools serving many students placed at risk. The program emphasizes research-based programs in prekindergarten and kindergarten; research-based curriculum and instruction in reading, writing, and language arts in grades 1-6; extensive professional development to help all teachers succeed with all children; one-to-one tutoring for students (primarily first graders having difficulties learning to read); and an active family support program. Table 1 summarizes the main elements of the model.

==========

Table 1

==========

Research on Success for All has found consistent positive effects of the program on measures of student achievement (Slavin et al., 1996; Slavin, Madden, Dolan, Wasik, Ross, & Smith, 1996; Madden, Slavin, Dolan, & Wasik, 1993; Ross, Smith, & Casey, 1996). Averaging data across 23 experimental and control schools in nine districts, Slavin et al. (1996) reported mean effect sizes of about +0.50 in every grade, 1-5, with grade equivalent differences on individually-administered tests increasing from almost three months at the end of first grade to more than a full grade equivalent in grade 5. Effect sizes in excess of +1.00 were found for the lowest-achieving 25% of students and for Spanish-dominant students taught in Spanish or English. Reductions in special education placements (Smith, Ross, & Casey, 1994) and substantial increases in the achievement of students already in special education (Ross, Smith, Casey, & Slavin, 1995) have also been found.

Since piloting the program in five inner-city Baltimore schools in 1987-89, Success for All has expanded substantially in use. As of spring, 1997, it is used by more than 475 schools in 31 states. Almost all are high-poverty, Title I schoolwide projects, but otherwise the schools vary

widely, from large cities to small towns to rural locations and Indian reservations. The program is primarily paid for from schools' existing Title I funds, often supplemented by state compensatory education, special education, bilingual, or other categorical funds. To support high-quality dissemination of Success for All, Johns Hopkins University and three regional training programs have developed a national network of full-time trainers, curriculum developers, and researchers, who are able to carry out continuing development and evaluation as well as to add more than 200 additional schools to the network each year.

Research Problem

Despite the success of Success for All in both research and dissemination, there are several key limitations of the approach. The most important is its cost. The program provides 20 person-days of training and follow-up to each school during the first implementation year, plus substantial instructional materials, which together cost a typical school of 500 students about $45,000 to $55,000 (plus about $15,000 per year for follow-up and replacement materials in the second and third years). Further, the model requires a full-time facilitator (an experienced teacher to ensure the quality of program implementation), plus several certified teachers and/or paraprofessionals to serve as tutors. Title I schoolwide projects with annual allocations of more than $150,000 can usually afford the program by reallocating existing funds and staff, but less well funded Title I or non-Title I schools have greater difficulty in doing so.*

At present, then, the application of Success for All is, as a practical matter, limited to high poverty schools with substantial categorical funding, or to other schools willing or able to recruit additional funds or engage in painful reallocation of non-categorical funds (see Odden, 1996, for a discussion of this).

---

* There is an irony here. The key issue is the level of Title I and other categorical funds, not overall per-pupil costs. High-poverty schools with high Title I but low per-pupil costs can typically afford the program, while low-poverty schools with high per-pupil costs cannot, because their funds are tied up in salaries, personnel, and other uses that are difficult to change.

Since the early years of research on Success for All, there has been a concern about evidence to support each of the major components of the model. If some components turn out to be unnecessary to the overall impact of the model, or if some components work well by themselves, without the others, then it might be possible to disseminate effective portions of the model to a much wider range of schools. In fact, an early study (Slavin, Madden, Karweit, Dolan, & Wasik, 1989) found that the reading program used in Success for All did have a positive impact on student achievement, though this impact was much smaller than that of the full model. Other comparisons (Slavin, Madden, Karweit, Dolan, & Wasik, 1992) compared a slimmed-down, "Chevrolet" version of Success for All, implemented entirely under existing Title I funds, to a "Cadillac" version in which significant additional funds were added to hire additional tutors, full-time rather than half-time facilitators, and family support staff. The differences between "Chevy" and "Cadillac" schools in student reading gains turned out to be relatively small for students in general, but quite large for students in the lowest 25% of their classes (who received most of the additional tutoring services).

Based in part on these findings and on practical experience in the field, the Success for All project established a high standard for resources needed to implement the program, including a full-time facilitator, several tutors, and 20 person-days of training, as noted earlier. This level of resourcing is still less than the "Cadillac" model, but more than the "Chevy;" increases since the late 1980's in overall Title I funding have made this strategy practicable, at least among the highest-poverty, highest-funded Title I schools.

Despite the findings of the earlier research, the question of the incremental effects of each of the major elements of Success for All has not yet been fully resolved. The earlier research involved a small number of schools, necessitating analysis of individual-level data, when school-level effects are clearly more appropriate for these school-level questions. To more conclusively answer questions about the effects on student outcomes of varying degrees of implementation of

the model, it was necessary to have a district able to implement alternative versions of the program in a large number of schools at the same time.

The Houston Project

The opportunity to do such a study came in 1994. In the summer of that year, the Houston Independent School District implemented a summer school program for first graders at risk of being held back. This program used reading materials and training from the Success for All project in 101 Houston elementary schools.

The summer school project was extremely successful--program students were retained at a rate of only 2.0% versus 8.6% of comparison students (HISD, 1995). Teachers also had very positive perceptions of Success For All-over 90% agreed with the philosophy of the program, thought the teaching methods were effective with their students, thought the model was flexible enough to adapt to the needs of their students, and felt confident in their ability to implement the model after training (HISD, 1995). As a result, HISD officials decided to offer to its highest-poverty elementary schools the opportunity to implement Success for All. However, as an experiment, schools could opt to implement just the reading program, reading plus tutoring, or the entire Success for All model. Fifty schools volunteered for one of these three groups. Twenty three matched comparison schools were then designated from among the pool of schools that did not take part in the summer school program.

Because the decision to implement the program came late in the summer, and because of the sheer scale of the implementation, training and materials for the project were not fully in place until well into the fall. In particular, the Spanish bilingual version of the program, used in 20 schools, did not get fully under way until spring, and therefore evaluation results for this subgroup were not used for this first cohort. However, a study of the much larger English-dominant sample found significant positive effects for all three program variations among schools that had fully implemented their treatments, although effects were much stronger among the schools

implementing the full model than among those implementing the reading program or reading and tutoring programs alone (Nunnery, 1995).

The present paper reports the results from the second implementation year in the Houston study, 1995-96. By this school year, bilingual as well as English programs in all schools had had time to implement their versions of the Success for All model, although there was considerable shifting of treatment categories after the initial assignment. Still, the quasi-experimental design of this study allows for comparisons to be made of the achievement effects of various configurations of treatment components, from full implementation to implementations lacking some elements felt to be essential to the model, especially full-time facilitators and certified teacher-tutors. The practical importance of the Houston experiment is profound. If it is possible to obtain student outcomes like those seen in earlier studies of Success for All with portions of the program or with significantly reduced resources, this would open up possibilities for implementation of these program elements among a far broader spectrum of schools, and perhaps reduce the program's cost and difficulty for its core constituency, high-poverty Title I schools. On the other hand, if the results only support the full model, as suggested by the first year results and the earlier component analyses (Slavin et al., 1992), this finding would reinforce the program's usual policy of refusing to compromise on the major elements of the approach.

Beyond the particulars of this one program, the findings of the Houston study could have important implications for the broader reform of high-poverty schools. It could provide some indication of the level of effort needed to make a substantial difference in the achievement of students in such schools. If improved curriculum and instruction is enough to bring about major achievement gains, then scaling up successful models may be comparatively easy. On the other hand, if full implementation of a program as complex as Success for All is necessary, then the consequences for educational policies point toward a substantial effort to fundamentally change the operating environments of high-poverty schools.

Methodology

Participants-English program. Two student cohorts participated in the study: (a) those who began first grade in 1995 (Cohort 1), and (b) those who began first grade in 1996 (Cohort 2). A pretest measure was administered to all Cohort 1 students (n=4,256). In 1996, posttests were given to approximately 10 Cohort 1 (2nd grade; n=595) and 10 Cohort 2 (1st grade; n=682) students from each school. The number of Cohort 1 students was somewhat reduced due to listwise deletion of cases for whom pretest data were not available. Complete data were available for a total of 46 SFA schools and 18 comparison schools. As a whole, SFA and comparison schools were similar in the average percentage of students eligible for free lunch (79% vs. 77%, respectively), although SFA schools had lower average percentages of Hispanic students (47% vs. 57%) and higher average mobility rates (53% vs. 30%). Comparison schools had slightly higher average pretest scores than SFA schools (M=49.4 vs. 48.6, SD=10.0).

Participants-Spanish bilingual program. Data were also collected for a cohort of Spanish dominant students who participated in the bilingual version of Success For All. The bilingual version of SFA was begun during the 1995-96 school year. However, the original intention was to begin during 1994-95, so the Spanish LAS was administered as a pretest to all entering first grade students in 1994-95 (n=1,682). Thus, the pretest could be used to group schools, but no pretest score was available at the individual level for the first bilingual cohort. All but one of the bilingual programs was housed in a school that was also implementing the English version of SFA. The final sample included 278 Spanish dominant first grade students in 20 SFA and 10 comparison schools

Reading measures. The Language Assessment Scales- Oral (LAS) was administered as a pretest to all English Cohort 1 students during fall of first grade. A battery of four reading posttests were individually administered to first graders in participating schools in April or May, 1995 (Cohort 1), and both second and first graders in 1996 (Cohort 1 and Cohort 2). The battery of reading tests included the *Word Attack*, *Word Identification*, and *Passage*

Comprehension subtests of the *Woodcock Reading Mastery Tests-Revised* (Woodcock, 1987), and the *Durrell Oral Reading Test* (Durrell & Catterson, 1980).

Transformation of pretest variable. Examination of spread and level plots of the pretest variance versus pretest means revealed a strong relationship between the natural logs of the means and the natural logs of the variances (slope = -3.5). A power transformation was used to induce homoscedasticity in the pretest covariate. The spread and level plot of the transformed variable showed the transformation to be successful in inducing both normality and homoscedasticity in LAS scores.

School composition variables. Six school composition variables were used to construct two orthogonal variables to represent aggregate background characteristics of students served by the school, and aggregate teacher characteristics: (a) average pretest LAS score, (b) percentage of students eligible for free or reduced-price lunch, (c) student mobility rate, (d) percentage of teachers with advanced degrees, (e) average years of experience of teachers in the school, and (f) teacher attendance rate. Because the sample of schools was positively skewed in terms of aggregate socioeconomic status of students, a normalizing transformation was performed on the free lunch and mobility rate variables by computing the arcsine of the square root of the percentage. The average pretest LAS was computed on the transformed pretest variable. Schools were classified as either majority-African American or majority-Hispanic according the percentage enrollment of students from each respective ethnic group.

School composition factor scores. A factor analysis using principal components extraction and a varimax rotation was performed on six school composition variables to generate a reduced set of independent variables representing school composition. For schools implementing the English language version of SFA, the analysis yielded two orthogonal factors that accounted for 55.5% of the variance in the set of scores: (a) teacher experience, upon which percentage of teachers with advanced degrees and average teacher experience loaded positively, and teacher

attendance rate loaded negatively; and (b) aggregate student background, upon which mobility rate and percentage eligible for free lunch loaded positively, and pretest mean loaded negatively.

The factor score representing student background was multiplied by -1, so high scores represent comparatively high pretest mean, low mobility rate, and low percentage eligible for free lunch. The factor scores were then converted to T-scores(M=50, SD=10)to aid interpretation. A comparison between English SFA and control schools showed that control schools had higher student background scores (M_c=52.29 versus M_{SFA}=48.97) and lower teacher characteristic scores (M_c =48.78 versus M_{SFA}=50.55). Schools were grouped into two strata (low, high) on the basis of a median split of the student background variable. However, it should be noted that "high" is a relative term. The average proportion of students qualifying for free lunch in the "low" group was 84.7%, while in the "high" group it was 71.2%, still very impoverished by national standards. The average pretest score of schools in the high group was 0.35 standard deviation units higher than the average for the low group.

Implementation categories. An implementation questionnaire survey was administered to principals or facilitators in all SFA Houston sites. A 100% response rate was obtained after 3 mail and 2 phone followups. The questionnaire solicited basic factual information about the program, such as the number and type of tutors, facilitator status (none, part time, full time), and whether the school implemented a family support team. An overall implementation score was computed by summing the standardized scores for facilitator status, family support, and number of certified tutors in the appropriate language per 100 students in grades 1 to 3. Note that the same school could receive different tutoring scores, depending on the number of tutors in Spanish and English. Schools were grouped into three categories (low implementation, medium implementation, and high implementation) based upon their implementation scores. Table 2 summarizes descriptive data on each category of schools.

==========

Table 2 here

==========

Description of English language program implementation levels. Across categories, schools provided about 2 tutors per 100 students in grades 1 to 3; however, medium-implementation and high-implementation implementation sites relied equally on certified and paraprofessional tutors, whereas low-implementation sites relied primarily on paraprofessionals. Low-implementation and medium-implementation schools usually had a part-time facilitator, whereas high-implementation sites were about equally likely to have either a full-time or part-time facilitator. Only 38% of low-implementation sites implemented the family support team component, whereas nearly all medium-implementation (87%) and high-implementation (93%) schools did so. Schools were similar across categories in student composition with respect to percentage eligible for free lunch and mobility rate, although low-implementation schools served a slightly higher average percentage of African American students (M=45%, versus about 35% for the other categories).

Treatment comparisons. For each cohort-year, multivariate analyses of variance (MANOVA's) were performed to test for overall treatment differences. In each MANOVA, scores from the four reading posttests were treated as dependent variables, while implementation level, ethnicity of student body (majority Hispanic or majority African American), and student background stratum were treated as independent variables. Because there were no high-implementation, high stratum, majority African American schools, the three-way interaction term was excluded from the model. For Cohort 1 analyses, the transformed pretest score was used as a covariate. Pillais trace was employed as the criterion of multivariate significance (Pedhazur, 1982, p.753). Follow-up univariate tests were conducted when the multivariate hypothesis tests suggested significant treatment effects.

Where univariate effects were significant, standardized residual scores were computed for each student via regression on the pretest variable. A one-way analysis of variance on the residual scores was then performed, using Scheffe's procedure to make all pairwise comparisons between implementation categories. For Cohort 1, effect size estimates were computed as the difference between mean standardized residual scores of a given SFA implementation level and the comparison mean. For Cohort 2, effect size estimates were computed as the standardized difference between posttest means.

Spanish bilingual program analyses. For the most part, the analyses of SFA bilingual program effects were conducted in the same manner as those of the English program. Spanish dominant students were given *Word Attack, Word Identification,* and *Passage Comprehension* posttest measures in Spanish. Implementation levels were constructed somewhat differently for bilingual programs. Only six schools had any certified tutoring in Spanish, so the ratio of paraprofessional tutors was also used to compute the overall implementation score. To retain adequate power and balance in the design, only two implementation categories were used (low-implementation, high-implementation). High-implementation schools exceeded low-implementation schools on all implementation variables. High-implementation schools were more likely to have family support teams (92% vs. 67%), and had higher tutoring rates for both certified tutors (0.23 per 100 students vs. 0.05) and paraprofessional tutors (2.54 vs. 0.88; see Table 2). High-implementation schools were more likely to have full-time facilitators (33% vs. 0%), and low-implementation schools were more likely to have no facilitator (25% vs. 0%). Spanish bilingual implementations differed from English language implementations in their reliance on paraprofessional tutors and lack of certified tutoring.

==========

Table 3 here

==========

Surprisingly, pretest means for Spanish bilingual students did not correlate with percentage eligible for free lunch or mobility rate. Thus, a composite student background score could not be constructed from these variables. Instead, a median split was performed on the pretest means to yield a two-level student background variable (low, high).

Results

English-dominant Cohort 1, 1996 (second graders). The two-way interaction effect between student racial majority status and implementation category was significant ($F_{12,1740}$=2.49, $p < .01$), as were the main effects for student racial majority status ($F_{4,578}$=3.17, $p = .01$) and implementation category ($F_{3,1740}$=2.17, p =.01). A graphical analysis was performed to determine whether the interaction was ordinal or disordinal by computing a composite residual achievement score and plotting the means for each interaction cell (Kennedy & Bush, 1995, p.268). As Figure 1 shows, the interaction was clearly disordinal.

==========

Figure 1 here

==========

Generally, higher implementation levels were associated with higher achievement. Within treatment levels, schools in which a majority of students were African American had higher scores than schools serving mostly Hispanic students. The disordinal interaction occurs because medium- and high- implementations had nearly identical effects on Hispanic students (M= +0.11 and +0.05, respectively), whereas there was a tremendous difference in medium-implementation and high-implementation effects in schools serving mostly African American students (M=-0.16 vs. +0.43). High-implementation, predominately African American schools were the only schools that substantially exceeded controls in terms of pretest-adjusted reading achievement, obtaining effect

size estimates from +0. 14 to +0.49 (see Table 3). High-implementation African American schools were the only schools in which, on average, students met or exceeded grade level on every test.

==========

Table 4 here

==========

English-dominant Cohort 2, 1996 (first graders). A significant multivariate interaction occurred between implementation level and socioeconomic strata ($F_{12,2004}$=1.82, p =.04). Followup univariate tests showed that the interaction effect was attributable to variation in *Word Attack* means ($F_{3,669}$=4.31, $p < .01$), but there was no significant interaction effect on the other measures. Control, low-implementation, and medium-implementation means showed an ordinal interaction pattern, where students scored higher in high stratum than low stratum schools, medium-implementation means exceeded comparison means, and low-implementation means were below comparison (see Table 5 and Figure 2). Among high-implementation schools, however, low stratum means were higher than high stratum means, suggesting that implementation made the largest difference in the most impoverished schools.

==========

Table 5 here

==========

==========

Figure 2 here

==========

Implementation category had a significant multivariate main effect ($F_{12,2004}$=3.78,

$p < .001$), which resulted from significant effects on all measures-*Oral Reading* ($F_{3,669}=7.68$, $p < .001$), *Passage Comprehension* ($F_{3,669}=7.28$, $p < .001$), *Word Attack* ($F_{3,669}= 10.08$, $P < .001$), *and Word Identification* ($F_{3,669}=9.32$, $p < .001$). Pairwise comparisons showed that comparison, medium-implementation, and high-implementation schools had significantly higher means than low-implementation schools on every measure. Medium-implementation schools had higher *Word Attack* means than comparison schools, while high-implementation schools had higher means than comparison on *Word Attack* and *Word Identification*.

Spanish dominant Cohort 1, 1996 (first graders). Multivariate tests showed that the interaction between implementation level and student background strata was significant ($F_{6,582}=2.37$, $p =.03$). Main effects for implementation level and strata were not significant. Followup univariate tests suggested that the multivariate effect was due to significant interactions effects on *Word Attack* ($F_{2,292}=3.73$, $p =.03$) and *Word Identification* ($F_{2,292} =5.16$, $p < .01$). As Table 6 and Figure 3 show, students in high-implementation schools exceeded comparison students in both low and high student background strata, whereas students in low-implementation schools scored below comparison in the low stratum and above comparison in the high stratum. Followup pairwise comparisons revealed that both low-implementation and high-implementation schools had significantly higher means than comparison schools in the high stratum, whereas all means were statistically equal in the low stratum.

==========

Table 6 here

==========

==========

Figure 3 here

==========

Discussion

Variations in implementation

As expected, implementations varied in terms of tutoring programs, facilitator status, and family support. In both Spanish and English programs, high implementation sites were much more likely to have a full-time facilitator. In English programs, low implementations provided an average of only 1 certified tutor per 200 students. Certified tutoring was infrequent in all bilingual implementations, but high implementation sites provided 3 times as many paraprofessional tutors per 100 students as low implementation sites (2.54 vs. 0.88). According to district policy, bilingual teachers who are reassigned to tutoring or facilitator roles lose their $3,000 annual bonus for teaching bilingual classrooms. This fact, coupled with a general shortage of bilingual teachers in the district, militated against high implementation levels of the bilingual program.

Achievement effects by implementation level: English version

Main effects of treatment levels. Although there were complex patterns of interaction between implementation level, majority ethnicity of student body, and student background status, some clear tendencies emerged. Low-level implementations were associated with achievement effects equal to or lower than comparison schools. High-level implementations generally produced better achievement than comparison schools, whereas medium-level implementations produced mixed results for the first cohort and positive results (compared to controls) in the second cohort. Overall, second-year results were more positive than first-year results, probably due to increased quality of implementation. This finding is consistent with previous research showing increasingly positive effects of SFA on each new cohort (Slavin et al., 1996).

Implementation by school composition interaction effects. Student achievement in extremely impoverished and predominately African American schools seemed far more sensitive to variations in implementation than in less impoverished or predominately Hispanic schools. Large

positive effects consistent with previous smaller scale SFA implementations were obtained with high level implementations in extremely impoverished schools and in schools serving predominately African American student populations. High and medium level implementations had nearly identical effects in moderately impoverished schools (average *ES* about +0.25) and in predominately Hispanic schools (average *ES* about +0.05). Medium level implementations exceeded comparisons regardless of socioeconomic stratum, but only high level implementations had positive effect sizes for predominately African American schools.

Students in low implementation sites consistently scored lower than students in comparison sites. In predominately African American schools, the deficit was small (average *ES*=-0.23) and primarily attributable to differences in decoding ability (Word Attack and Word Identification). In predominately Hispanic schools, low implementation was associated with larger deficits (average *ES* = -0.43) across all measures.

Achievement effects by implementation level: Spanish bilingual version.

Implementation level effects varied by student background stratum in Spanish bilingual programs. Regardless of students' aggregate prior achievement, medium level implementations produced better effects than comparison schools. Low-implementation schools performed worse than comparison in schools serving larger proportions of low achieving children, but better in schools serving students with relatively high levels of language development prior to entering school.

Conclusions

Clearly, schools that implemented variations of SFA that were closest to the original model achieved better results. The combined influence of certified tutoring, a full-time school-level facilitator, and the family support team were necessary to enhance student achievement in the most impoverished schools. Programs with part-time facilitators, certified tutoring, and a family

support team seemed to produce positive effects in moderately impoverished schools, but not in extremely impoverished schools. Schools that did not make the basic organizational changes required to provide family support and certified tutoring performed worse than comparison schools.

Providing a full-time program facilitator seemed an essential program component in extremely high-poverty schools serving predominately African American student populations. Students in moderate-poverty or predominately Hispanic schools seemed to achieve about equally well with full-time or part-time facilitators, given sufficient certified tutoring and a Family Support Team. However, many of the facilitators in Hispanic schools were simultaneously charged with facilitating the bilingual program. Whether full- or part-time, facilitators in Hispanic schools were thus spread more thinly than those in schools serving mostly African American students. This may account for the lack of differentiation of high- and medium-implementation effects in Hispanic schools.

Schools that provided certified tutoring clearly achieved better results in the English language program. All schools averaged about 2 tutors per 100 students regardless of implementation category, but the group that relied mostly on paraprofessionals had negative effect size estimates in every circumstance. It is tempting to infer that high implementation sites achieved better results because of the direct effects of certified tutoring. However, the relatively small number of students sampled at each school makes it unlikely this effect is due to effects only on individuals who received tutoring. Rather, provision of certified tutoring may be a proxy measure indicating the school's commitment to the program. By using paraprofessionals, low implementation schools avoid altering staff roles or resource allocation patterns. Conversely, school staffs that make the effort to provide certified tutoring may be more likely to make other necessary organizational changes.

The findings reported here suggest that schools, especially those serving the most impoverished students, are best advised to undertake implementation of the whole program rather

than implement parts. However, the elements of the program may be multiplicative in effect, rather than additive, such that large differences in student achievement are obtained only when all elements are present. This parallels results of a study of variation in component configurations of the Reading Recovery tutoring model, which indicated that only the full model was effective (Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994).

School-level commitment to SFA, as indicated by a willingness to undertake fundamental restructuring of staff roles and resource allocation, seems predictive of ultimate success in terms of student achievement. The district undoubtedly solicited involvement from a larger number of schools by reducing, in effect, the amount of change required. However, by attempting to implement a reduced form of the program that leaves out key program components, schools may ultimately be less successful in achieving any meaningful change that influences student achievement. A perennial finding in school change research is that reforms are more successful if they are complex and comprehensive (e.g., Berman & McLaughlin, 1975). Schools attempting less comprehensive change are more likely to adopt programs in a pro forma manner, without seriously altering the organizational status quo.

It is important to note that even before the data presented here were known, HISD officials and Johns Hopkins trainers were aware that the less complete implementations of Success for All were not only working poorly, but were also undermining the quality of implementation in the full model schools. The full model schools had difficulty holding to all program elements because their staffs were aware that other schools in the district were implementing reduced versions, which were easier and less expensive to implement. As a result, at the end of the 1995-96 school year, all schools in the Success for All treatments were required by the district to choose whether to implement the entire program or to drop out. All but one school chose to continue, and as a result, implementations during the 1996-97 school year, under way at this writing, are significantly better and more complete than was the case previously.

Implications for School Reform

The implications of the findings of the Houston experiment for school reform in general are clear. Consistent with the findings of the Ohio statewide study of Reading Recovery (Pinnell et al., 1994), the present findings suggest that fundamental comprehensive changes in teachers' daily practices and in school organization are needed to bring about major changes in the reading success of students placed at risk, especially those in the most impoverished schools. These findings dispel the belief that bringing all children to adequate levels of reading performance will be easy or inexpensive. The good news, from this study as well as many other studies of Success for All, is that we can reliably make a profound difference in the reading performance of students in our most disadvantaged schools. Recall that the high-implementation African American schools in this study had students scoring above national norms. However, to accomplish these results on a broad scale may take nothing less than the full efforts and full costs required by the Success for All model or comparable alternatives.

References

Berman, P., & McLaughlin, M. W. (1975). *Federal programs supporting educational change, Vol. IV.- The findings in review.* Santa Monica, CA: The Rand Corporation.

Durrell, D. D., & Catterson, J. H. (1980). *Durrell Analysis of Reading Difficulty, 3rd Edition.* The Psychological Corporation.

HISD (1995). *Retention Reduction Summer 1995 evaluation report.* Houston, TX: Department of Research and Evaluation, Houston Independent School District.

Kennedy, J., & Bush, A. J. (1985). *An introduction to the design and analysis of experiments in behavioral research.* Lanham, MD: University Press of America.

Madden, N.A., Slavin, R.E., Karweit, N.L., Dolan, L., & Wasik, B.A. (1990, April). *Success For All: Effects of variations in duration and resources of a schoolwide elementary restructuring program.* Paper presented at the annual convention of the American Educational Research Association, Boston.

Madden, N. A., Slavin, R.E., Karweit, N. L., Dolan, L. J., & Wasik, B.A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30,* 123-148.

Nunnery, J. A. (1995). *An assessment of Success For All program component configuration effects on the reading achievement of at-risk first grade students.* Unpublished doctoral dissertation, The University of Memphis.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction.* Fort Worth, TX: Holt, Rinehart, and Winston, Inc.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high risk first graders. *Reading Research Quarterly, 29,* 8-38.

Slavin, R. E., Madden, N. A., Dolan, L., & Wasik, B. A. (1996). *Every child, every school: Success for All.* Newbury Park, CA: Corwin.

Slavin, R.E., Madden, N., Dolan, L., Wasik, B., Ross, S., Smith, L., & Dianda, M. (1996). Success for All: A summary of research. *Journal of Education for Students Placed At Risk, 1,* 41-76.

Smith, L.J., Ross, S.M., & Casey, J.P. (1994). *Special education analyses for Success for All in four cities.* Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Smith, L., Ross, S., & Casey, J. (1994). *Special education analyses for Success for All in four cities.* Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Ross, S.M., Smith, L.J., Casey, J., & Slavin, R.E. (1995). Increasing the academic success of disadvantaged children: An examination of alternative early intervention programs. *American Educational Research Journal, 32,* 773-800.

Woodcock, R. (1987). *Woodcock Reading Mastery Tests-Revised.* Circle Pines, MN: American Guidance Service, Inc.

Woodcock, R. (1981). *Woodcock Language Proficiency Battery: Spanish Form.* Chicago, IL: The Riverside Publishing Company.

Table 1

Major Elements of Success for All

Success for All is a schoolwide program for students in grades pre-K to five which organizes resources to attempt to ensure that virtually every student will reach the third grade on time with adequate basic skills and build on this basis throughout the elementary grades, that no student will be allowed to "fall between the cracks." The main elements of the program are as follows:

1. *Tutors.* In grades 1-3, specially trained certified teachers work one-to-one with any students who are failing to keep up with their classmates in reading. Tutorial instruction is closely coordinated with regular classroom instruction. It takes place 20 minutes daily during times other than reading periods.

2. *A School-Wide Curriculum.* During reading periods, students are regrouped across age lines so that each reading class contains students all at one reading level. Use of tutors as reading teachers during reading time reduces the size of most reading classes to about 20. The reading program in grades K- 1 emphasizes language and comprehension skills, sound blending, and use of shared stories that students read to one another in pairs. The shared stories combine teacher-read material with phonetically regular student material to teach decoding and comprehension in the context of meaningful, engaging stories. In grades 2-5, students use novels or basals but not workbooks. This program emphasizes cooperative learning activities built around partner reading, identification of characters, settings, problems, and problem solutions in narratives, story summarization, writing, and direct instruction in reading comprehension skills. At all levels, students are required to read books of their own choice for twenty minutes at home each evening. Classroom libraries of trade books are provided for this purpose. Beginning in the second year of implementation, cooperative learning programs in writing/language arts are introduced in grades K-5.

3. *Preschool and Kindergarten.* The preschool and kindergarten programs in Success for All emphasize language development, readiness, and self-concept. Preschools and kindergartens use thematic units, Peabody Language Development Kits, and a program called Story Telling and Retelling (STaR).

4. *Eight-Week Assessments.* Students in grades K-3 are assessed every eight weeks to determine whether they are making adequate progress in reading. This information is used to suggest alternate teaching strategies in the regular classroom, changes in reading group placement, provision of tutoring services, or other means of meeting students' needs.

5. *Family Support Team.* A family support team works in each school to help support parents in ensuring the success of their children, focusing on parent education, parent involvement, attendance, and student behavior. This team is composed of existing or additional staff such as parent liaisons, social workers, counselors, and vice principals.

6. *Facilitator.* A program facilitator works with teachers to help them implement the reading program, manages the eight week assessments, assists the family support team, makes sure that all staff are communicating with each other, and helps the staff as a whole make certain that every child is making adequate progress.

Table 2

Selected Program Characteristic Means by Implementation Category—English Language Version

Implement Category	Certified Tutoring[1]	Aide Tutoring	Facilitator[2]	FST[3]	% Hispanic	% African American	% Free Lunch	Mobility
Comparison					59.8	32.2	76.7	32.0
SFA/Low	0.49	1.41	1.00	0.38	49.3	45.5	79.9	37.1
SFA/Medium	0.97	1.03	0.87	0.87	52.7	37.5	79.5	34.8
SFA/High	1.19	0.97	1.47	0.93	54.5	35.3	76.2	30.4

[1]Number of English language tutors per 100 English-dominant students in grades 1 to 3. [2]Program facilitator (0=none, 1=part-time, 2=full-time). [3]Family Support Team (0=none, 1=implemented).

BEST COPY AVAILABLE

Table 3

Selected Program Characteristic Means by Implementation Category—Spanish Bilingual Version

Implement. Category	Certified Tutoring[1]	Aide Tutoring	Facilitator[2]	FST[3]	LAS *M*[4]	% Free Lunch	Mobility
Comparison					71.0	79.6	30.9
SFA/Low	0.05	0.88	0.75	0.67	69.2	80.8	36.2
SFA/High	0.24	2.54	1.33	0.92	72.4	84.5	38.7

[1]Number of Spanish language tutors per 100 Spanish-dominant students in grades 1 to 3. [2]Program facilitator (0=none, 1=part-time, 2=full-time). [3]Family Support Team (0=none, 1=implemented). [4]Average pretest mean score on *Language Assessment Scales*.

Table 4

Adjusted Mean Grade Equivalent Scores (GE') and Effect Size Estimates by Implementation Level and Majority Ethnic Group of School. 1996 English Cohort 1 (Second Grade)

Level	Oral Reading		Passage Comprehension		Word Attack		Word Identification	
	GE'	ES	GE'	ES	GE'	ES	GE'	ES
Low-implementation								
African American	2.8	-0.04	2.6	-0.14	2.4	-0.40	2.8	-0.32
Hispanic	2.5	-0.39	2.1	-0.61	2.3	-0.24	2.4	-0.45
Medium-implementation								
African American	2.8	-0.13	2.4	-0.36	2.4	-0.37	2.6	-0.32
Hispanic	2.8	+0.03	2.6	+0.02	2.9	+0.15	3.1	+0.11
High-implementation								
African American	2.9	+0.49	2.9	+0.18	3.1	+0.14	3.4	+0.22
Hispanic	2.7	+0.15	2.6	-0.01	2.6	+0.01	2.9	-0.03
Comparison								
African American	2.8		2.7		2.9		3.2	
Hispanic	2.8		2.6		2.6		2.9	

Table 5

Mean Grade Equivalent Scores (GE) and Effect Size Estimates by Implementation Level and Socioeconomic Stratum. 1996 English Cohort 2 (First Grade) Results

Level	Oral Reading		Passage Comprehension		Word Attack		Word Identification	
	GE	ES	GE	ES	GE	ES	GE	ES
Low-implementation								
Low	1.5	-0.29	1.6	-0.34	1.6	-0.18	1.6	-0.34
High	1.7	-0.31	1.7	-0.29	1.9	-0.07	1.8	-0.20
Medium-implementation								
Low	2.0	+0.23	1.9	+0.23	1.9	+0.36	1.9	+0.13
High	2.1	+0.14	1.9	+0.18	2.1	+0.25	2.1	+0.26
High-implementation								
Low	2.0	+0.33	1.9	+0.34	2.3	+0.73	2.3	+0.55
High	2.1	+0.10	1.9	+0.05	2.1	+0.21	2.1	+0.13
Comparison								
Low	1.8		1.7		1.7		1.8	
High	2.0		1.8		1.9		1.9	

Table 6

Mean Grade Equivalent Scores (GE) and Effect Size Estimates by Implementation Level and Student Background Stratum. Spanish Bilingual Program. 1996 First Grade

Level	Passage Comprehension		Word Attack		Word Identification	
	GE	ES	GE	ES	GE	ES
Low-implementation						
Low	2.5	-0.25	2.2	-0.20	2.5	-0.18
High	2.8	+0.29	3.9	+0.58	3.3	+0.67
Medium-implementation						
Low	2.9	-0.02	3.1	+0.10	3.0	+0.15
High	3.6	+0.51	3.7	+0.52	2.8	+0.28
Comparison						
Low	2.9		2.8		2.8	
High	2.5		2.2		2.4	

Note. Grade equivalent scores seem implausibly high. At the time of writing, no explanation for this result was available. This suggests either (a) a consistent positive bias in the raw score means, or (b) an inappropriate norm reference group for translating raw scores into grade equivalent scores.

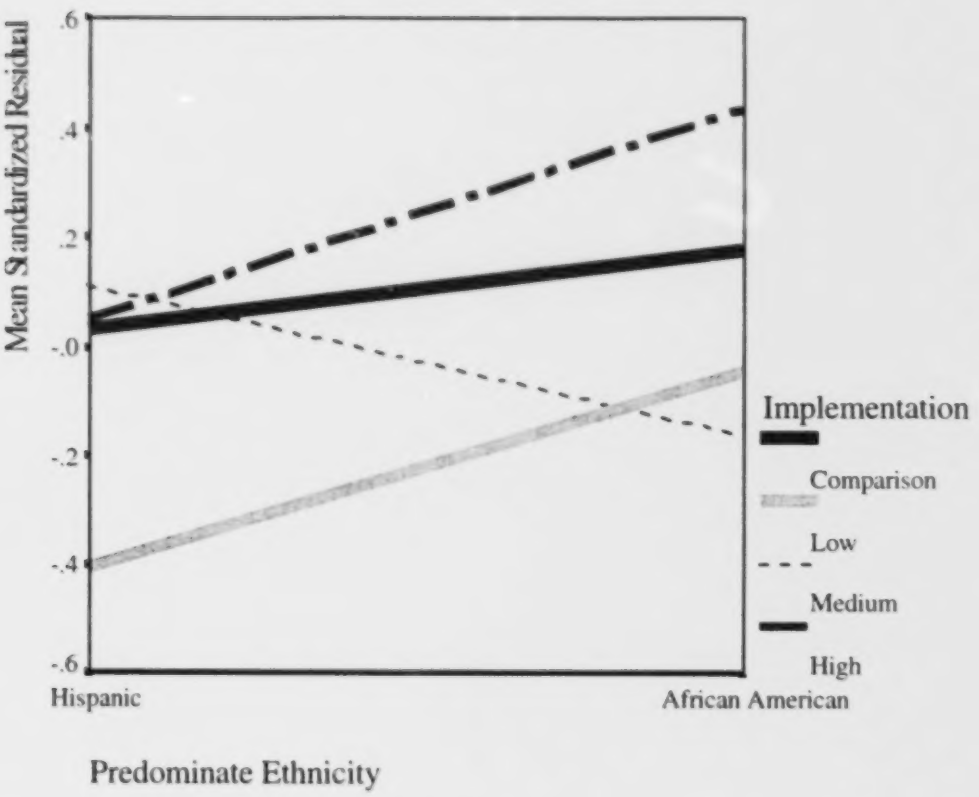

Figure 1. Composite Standardized Residual Achievement Scores by Implementation Category and Ethnicity of Majority Enrollment, English Cohort 1, 1996

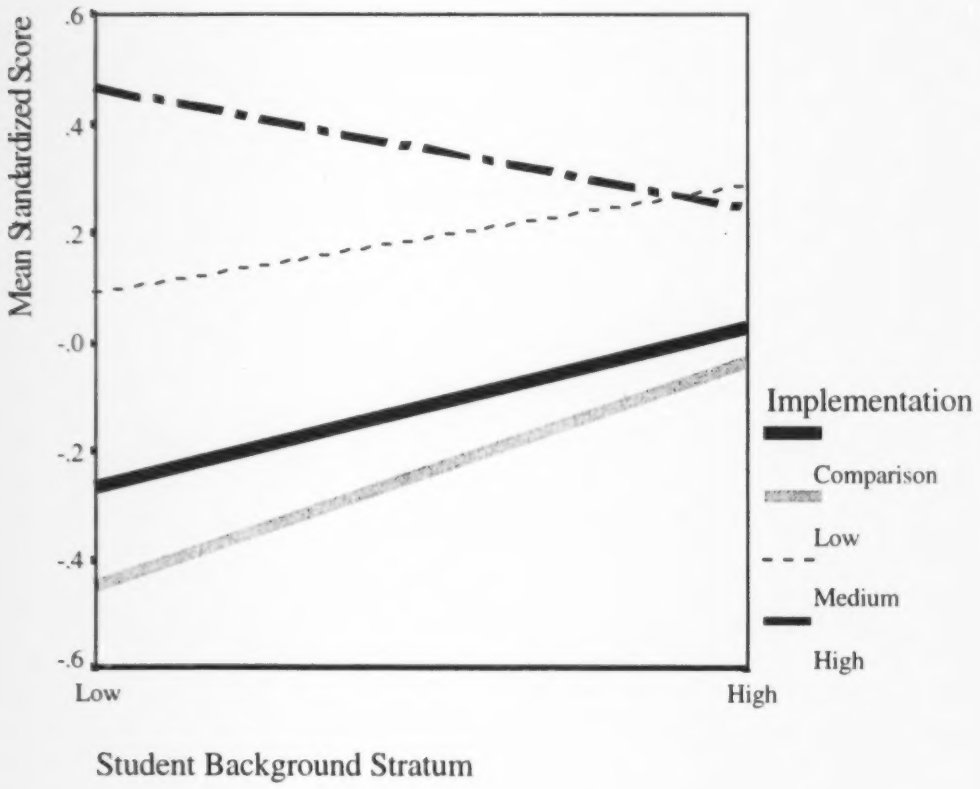

Figure 2. Standardized Word Attack Means by Implementation Category and Socioeconomic Stratum, English Cohort 2, 1996

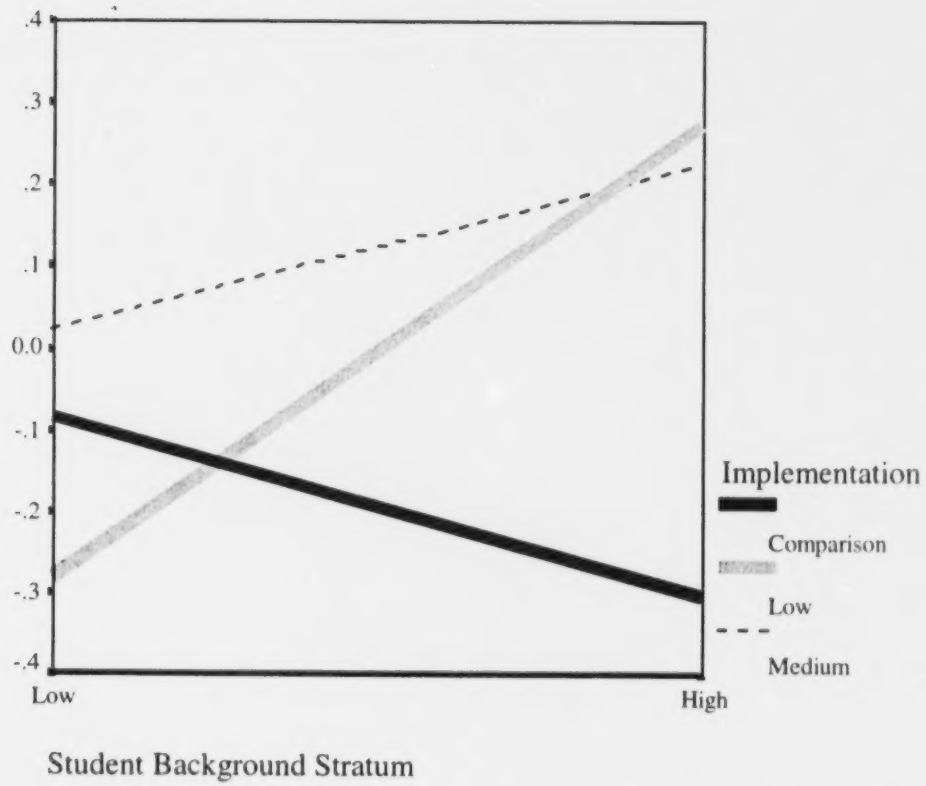

Figure 3. Word Attack Standardized Score Means by Implementation Level and Student Background Stratum, Spanish Bilingual Cohort 1 (First Grade), 1996.